



Europäisches Patentamt
European Patent Office
Office européen des brevets



Publication number: **0 625 572 A1**

(12)

EUROPEAN PATENT APPLICATION
published in accordance with Art.
158(3) EPC

(21) Application number: **93921061.3**

(51) Int. Cl.⁵: **C12N 15/10**

(22) Date of filing: **22.09.93**

(86) International application number:
PCT/JP93/01359

(87) International publication number:
WO 94/08001 (14.04.94 94/09)

(30) Priority: **25.09.92 JP 280932/92**
03.03.93 JP 67589/93

(43) Date of publication of application:
23.11.94 Bulletin 94/47

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC
NL PT SE

(71) Applicant: **THE KANAGAWA ACADEMY OF**
SCIENCE AND TECHNOLOGY FOUNDATION
3-2-1, Sakado, Takatsuku,
Kawasaki-shi
Kanagawa 213 (JP)
Applicant: **KATO, Seishi**
1-9-2-304, Minamidai,
Sagamihara-shi
Kanagawa 228 (JP)

Applicant: **SEKINE, Shingo**
4-4-1, Nishioonuma,
Sagamihara-shi
Kanagawa 229 (JP)

(72) Inventor: **KATO, Seishi 1-9-2-304, Minamidai**
Sagamihara-shi
Kanagawa 228 (JP)
Inventor: **SEKINE, Shingo 4-4-1, Nishioonuma**
Sagamihara-shi
Kanagawa 229 (JP)

(74) Representative: **Greaves, Carol Pauline et al**
Mewburn Ellis
York House
23 Kingsway
London WC2B 6HP (GB)

(54) **PROCESS FOR SYNTHESIZING COMPLETE-LENGTH cDNA, PROCESS FOR PRODUCING**
INTERMEDIATE THEREFOR, AND PROCESS FOR PRODUCING RECOMBINANT VECTOR CONTAINING
COMPLETE-LENGTH cDNA.

(57) The invention discloses a method of providing a cDNA containing the whole of the information on the primary structure of a protein by selectively synthesizing only the complete-length cDNA that contains a sequence beginning with the cap site of a mRNA. The process for producing an intermediate for the synthesis of the complete-length cDNA comprises the step of treating a mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped incomplete-length mRNA, the step of decapping from the 5' end of a capped complete-length mRNA,

and the step of linking the 5'-end phosphate group formed in the above step to either the DNA oligonucleotide represented by the following general formula (I) or a DNA-RNA chimera oligonucleotide by means of a T4RNA ligase to selectively add the DNA oligonucleotide or DNA-RNA chimera oligonucleotide having an arbitrary sequence to the 5' end of the complete-length mRNA: 5'-dN₁-dN₂-...-dN_m-N₁N₂-...-N_n-3', wherein dN represents a deoxyribonucleotide selected from among dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from among AMP, CMP, GMP and UMP; "-

EP 0 625 572 A1

" represents a phosphoric ester linkage; m represents an integer of 1 or above; and n represents an integer of 0 or above. The process for synthesizing a complete-length cDNA comprises linking a double-stranded DNA primer having a dT tail by annealing to the poly (A) tail of the 3' end of the complete-

length mRNA having the 5' end to which the DNA oligonucleotide or DNA-RNA chimera oligonucleotide have been added as described above, and then synthesizing a single-stranded cDNA complementary to the complete-length mRNA by means of a reverse transcriptase.

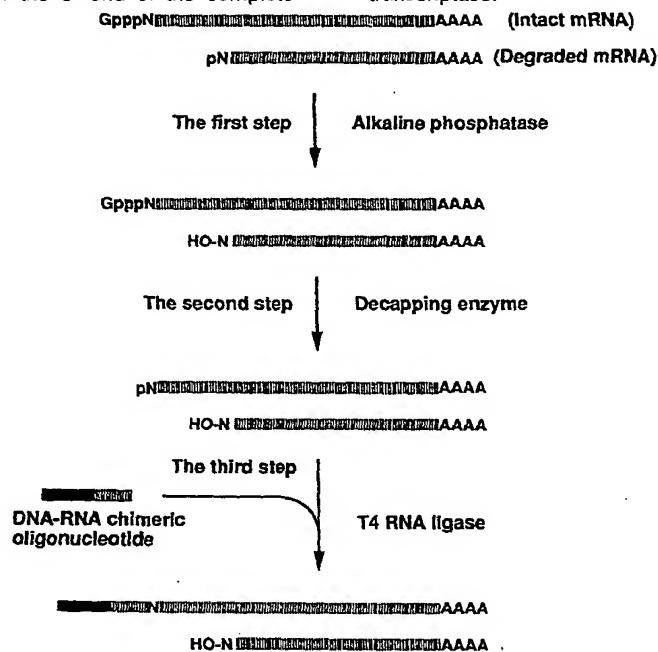


Fig.1

Technical Field

The present invention relates to a process for synthesizing a full-length cDNA from mRNA. This invention enables screening of proteins useful in industry and mass production of the proteins.

Background Art

Bio-active proteins produced by cells have been widely utilized in industry as materials for medicines, diagnosis, biosensors, bio-reactors and so on. The progress of gene technology has facilitated the discovery of these proteins and enabled us to produce a large amount of the proteins. The fundamental technology is a cDNA cloning technique.

Information on the amino acid sequence of a protein is encoded by mRNA. If the mRNA is converted to a corresponding DNA, i.e., a complementary DNA (cDNA), the cDNA can be used for determining the primary structure of the protein and producing a large amount of the protein. Therefore, there have been developed various cDNA cloning techniques which include isolating a mRNA from cells, synthesizing cDNAs from the mRNA, and isolating a cDNA encoding a target protein.

The most important requisite for cDNA cloning is to synthesize a cDNA containing an entire protein coding region encoded by an intact mRNA. The intact mRNA contains a characteristic structure at the 5' terminal, so called a cap. The cDNA containing an entire sequence involving from the nucleotide at which the cap is added is called a "full-length cDNA". If a full-length cDNA is synthesized, we can know the whole information on the primary structure of a protein and also use it for the mass production of the protein encoded by the cDNA. According to the conventional terminology, a "full-length cDNA" have been used as a cDNA containing an entire coding region of the protein even if the cDNA contains no cap site sequence. In this invention, however, a "full-length cDNA" is defined only as a cDNA containing a nucleotide sequence starting from the cap site.

The Gubler-Hoffman method [Gene 25:263-269(1983)], which is most widely used for cDNA synthesis, could not give a full-length cDNA, because deletion at the terminals of cDNA occurs. On the other hand, the Okayama Berg method [Mol.Cell.Biol. 2:161-170(1982)] is known to give a full-length cDNA at high efficiency. The characteristic of this method is to add a dC tail to the 3' end of the first strand cDNA synthesized from mRNA. This method, however, does not always give a full-length cDNA, because dC tailing also occurs to the truncated cDNA generated by unexpected termina-

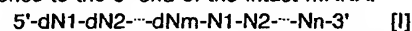
tion of cDNA synthesis. The truncated cDNAs originated from the degraded mRNAs are also synthesized.

Disclosure of the Invention

It is therefore an object of the present invention to provide a method for selectively synthesizing a full-length cDNA containing a nucleotide sequence starting from a cap site of mRNA, and a cDNA containing whole information on the primary structure of a protein. Furthermore another object of this invention is to provide a method to prepare an intermediate for synthesizing the full-length cDNA and to prepare a recombinant vector containing the full-length cDNA.

The inventors succeeded in synthesizing a full-length cDNA at high efficiency from mRNA whose cap structure was replaced by a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide.

The invention provides a process for producing an intermediate for the synthesis of a full-length cDNA, which comprises a step of treating mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped degraded mRNA, a step of decapping from the 5' end of a capped intact mRNA, and a step of ligating either a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide represented by the following general formula [I] to the 5' end phosphate group formed in the above step by the action of T4 RNA ligase, thereby selectively adding either the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide having an arbitrary sequence to the 5' end of the intact mRNA.



wherein dN represents a deoxyribonucleotide selected from among dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from among AMP, CMP, GMP, and UMP; "-" represents a phosphoric ester linkage; m represents an integer of 1 or above; n represents an integer of 0 or above.

Furthermore, the invention provides a process for synthesizing a full-length cDNA, which comprises linking a double-stranded DNA primer having a dT tail by annealing to a poly(A) tail of the 3' end of the intact mRNA having the DNA oligonucleotide or the DNA-RNA chimeric nucleotide added at the 5' end as described above, and then synthesizing the first strand cDNA complementary to the intact mRNA by means of a reverse transcriptase.

Furthermore, the invention provides a process for preparing a recombinant vector containing a full-length cDNA, which comprises the steps involved in the process for synthesizing a full-length

cDNA described above in which the sequence of the DNA oligonucleotide or the DNA-RNA oligonucleotide contains at least one site recognized by a restriction enzyme RE1 and the double-stranded DNA primer contains the same, the step of digesting with a restriction enzyme RE1 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by above method, the step of circularizing the resultant material by self-ligation, and the step of converting the RNA in the circularized recombinant vector to DNA.

The invention enables sure synthesis of a full-length cDNA which has been difficult to be synthesized according to the conventional methods. Since a cDNA clone obtained from the cDNA library prepared according to the present method is sure to contain the whole information on the primary structure of the protein, the obtained clone can be used immediately to produce the encoded protein.

Brief Description of the Drawings

Figure 1 shows a process to selectively ligate a DNA-RNA chimeric oligonucleotide to an intact mRNA possessing a cap; and

Figure 2 shows a process to synthesize cDNA from a synthetic oligonucleotide-ligated mRNA.

Best Mode for Carrying Out the Invention

This invention is concerned with a method for synthesizing a full-length cDNA, including two processes. The first process consists of ligation of a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide to an intact mRNA possessing a cap and the second process consists of cDNA synthesis from the synthetic oligonucleotide-capped mRNA and production of a recombinant vector containing the cDNA. Each step in the first process is shown in Figure 1. All steps in the first and second processes can be done using the conventional methods in this field.

The First Step

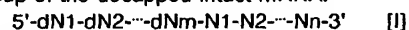
Usually mRNAs extracted from cells contain degraded ones possessing no cap structure. This step is carried out to remove the phosphate group of the degraded mRNA by means of an enzyme such as an alkaline phosphatase. Since an intact mRNA has a cap structure at the 5' end, the phosphatase does not affect the intact mRNA. The 5' end of a degraded mRNA possessing no cap structure is converted to a hydroxy group by selective dephosphorylation.

The Second Step

The cap structure of the intact mRNA is removed by a decapping enzyme to generate one phosphate group at the 5' end. The tobacco acid pyrophosphatase or T4 polynucleotide kinase can be exemplified as the decapping enzyme.

The Third Step

The DNA oligonucleotide or DNA-RNA chimeric oligonucleotide described by the following formula [I] is ligated to the 5'-terminal phosphate group of the decapped intact mRNA.



wherein dN represents a deoxyribonucleotide selected from among dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from among AMP, CMP, GMP, and UMP; "-" represents a phosphoric ester linkage; m represents an integer of 1 or above; n represents an integer of 0 or above.

In order to facilitate the followed preparation of a recombinant vector containing the cDNA, the sequence of the deoxyribonucleotide preferably contains at least one site to be recognized by a restriction enzyme RE1. Any restriction enzyme RE1 which does not cut the DNA-RNA hybrid can be used, and one generating the protruding site is preferred. In this step the 5' end of the intact mRNA, which has a phosphate group generated at the above second step, can incorporate with a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide. On the other hand, the 5' end of the degraded mRNA cannot incorporate with it because of the hydroxyl group generated at the 5' end by dephosphorylation in the first step. Thus the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide is selectively added to the 5' end of the intact mRNA. Since T4 RNA ligase catalyzes the ligation between RNA strands more efficiently than between RNA and DNA strands, the use of a DNA-RNA chimeric oligonucleotide is preferable.

The DNA oligonucleotide or DNA-RNA chimeric oligonucleotide used here can be chemically synthesized in a tube or with an automatic DNA/RNA synthesizer. There have been no reports on the DNA-RNA chimeric oligonucleotide in spite of its easy synthesis, so that it is provided by this invention for the first time.

It is preferable to use excess molar ratio of the oligonucleotide to mRNA in the ligation reaction between the oligonucleotide and mRNA.

Each step of the second process is shown in Figure 2.

The First Step

The mRNA capped by a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide, which was prepared in the first process, is annealed with a double-stranded DNA primer containing a tail of n (dT)s at the 3' end. Then the first strand cDNA is synthesized using reverse transcriptase. The double-stranded DNA primer can be selected from an oligonucleotide, any double-stranded DNA fragment or any vector primer (for example, a plasmid DNA containing a replication origin of E.coli). The oligonucleotide primer can be chemically synthesized. In this case, n is preferably from 10 to 30. The latter two primers can be prepared by cutting the plasmid vector with a restriction enzyme generating a protruding 3' end and adding a dT-tail using terminal deoxynucleotidyl transferase. In this case, n is preferably from 50 to 70.

The double-stranded DNA primer preferably possesses at least one site for restriction enzyme RE2 near the end to which the dT-tail is not added for easy preparation of a recombinant vector containing a cDNA. Any restriction enzyme RE2, preferably one generating a protruding end by digestion, can be used as long as it does not cut a mRNA-cDNA hybrid. The RE1 and RE2 should be identical when a vector primer is used, although they could be different when an oligonucleotide primer or a DNA fragment primer is used.

The synthesis of the first strand cDNA complementary to a intact mRNA is completed by the first step described above. The following second, third, and fourth steps are those for preparing a recombinant vector containing the synthesized cDNA.

The Second Step

When two restriction sites for RE1 and RE2 are used, the synthesized product containing the first strand cDNA is digested with both RE1 and RE2 to generate the corresponding restriction end at both ends of the double-stranded DNA parts. When no restriction sites exist in these parts, any restriction end can be prepared by adding linkers to the ends with T4 DNA ligase and digesting them with the corresponding restriction enzymes according to the conventional method. The blunt ends can be used without digestion with restriction enzymes.

The Third Step

This is a step for inserting the obtained cDNA into an appropriate cloning vector for E.coli. The procedure depends on the used primer. When an oligonucleotide primer or a DNA fragment primer is used, the cDNA fragment prepared in the second

step is ligated to any cloning vector (including plasmid, cosmid, phage and so on) for E.coli after digesting it with restriction enzymes RE1 and RE2. If a linker is used for preparing the restriction ends in the second step, the vector should be digested with the restriction enzymes corresponding to the site designed in the linker, or the corresponding linker should be added to the ends of the vector prior to the ligation to the cDNA fragment. In either cases, the ligation between blunt ends can be carried out without digestion with restriction enzymes.

The Fourth Step

The RNA strand in the vector is replaced by a DNA strand using RNase H, E.coli DNA polymerase I, and E.coli DNA ligase according to the conventional method.

Above procedure allows to prepare a recombinant vector containing a cDNA complementary to an intact mRNA.

The recombinant vector prepared by the above procedure can be used to transform host cells such as E.coli. The plasmid, cosmid, and phage etc. prepared from the transformants contain the full-length cDNA. Thus these processes give a clone containing a full-length cDNA.

The present invention will now be described by way of examples, but various changes and modifications can be made without departing from the invention.

REFERENCE EXAMPLES

(1) Preparation of mRNA

A model mRNA was prepared from the cDNA encoding human ribosomal phosphoprotein P1 (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292). Five µg of the cDNA plasmid was digested with NotI and then used as a template for in vitro transcription using a kit purchased from Boehringer Mannheim. By adding m7G(5')ppp(5')G in the reaction mixture, the cap was added to the 5' end of the transcribed RNA. When the product was analyzed by formamide-containing agarose gel electrophoresis, a single band of c.a. 600 nucleotides was shown on the gel. This was used as a model mRNA for cDNA synthesis. This model mRNA is composed of the vector-derived sequence starting from the cap, the sequence derived from the mRNA encoding human ribosomal phosphoprotein P1, and about 100 nucleotides of poly(A) tail. Not all mRNAs synthesized above contain the cap at the 5' end.

(2) Synthesis of DNA-RNA chimeric oligonucleotide

The oligonucleotide described by the following formula was synthesized with a DNA synthesizer (Applied Biosystems, model 392) using the phosphoramidite method.

5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-G-G-A-3'

Reagents for RNA synthesis and DNA synthesis were purchased from Peninsula and ABI, respectively. First the adenine-CPG column was set on the synthesizer. Then two ribonucleotides (G) followed by the other deoxyribonucleotides were added in order of above sequence under the reaction conditions described in the protocol. Finally the synthesized oligomer was cut in an ammonia-ethanol (3:1) solution. The protected groups were removed by treating with THF at 55°C for 10 h. The product was purified on polyacrylamide gel electrophoresis.

(3) Synthesis of DNA oligonucleotide

The DNA oligonucleotide described by the following formula was synthesized with a DNA synthesizer (Applied Biosystems, model 392) using the phosphoramidite method.

5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-dG-dG-dA-3'

The reaction conditions described in the protocol were used. The product was purified on polyacrylamide gel electrophoresis according to the conventional method.

(4) Preparation of the DNA fragment primer

One hundred µg of the pKA1 vector primer possessing a tail of c.a. 60 (dT)s (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292) was digested with 100 units of the restriction enzyme NdeI. The 500 bp fragment containing a (dT) tail was isolated on 1.5 % agarose gel electrophoresis. The resultant fragment was used as a DNA fragment primer.

EXAMPLE 1

One hundred µg of the model mRNA encoding human ribosomal phosphoprotein P1, which was prepared in above REFERENTIAL EXAMPLE, was dissolved in 100 mM Tris-HCl (pH 8.0). After adding one unit of RNase-free bacterial alkaline phosphatase (Takara Shuzo), the reaction mixture was incubated at 37°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in a solution containing 50 mM sodium acetate (pH 6.0), 1 mM EDTA, 0.1 % mercaptoethanol, and 0.01% Triton

X-100. After adding one unit of tobacco acid phosphatase (Epicentre Technologies), the reaction mixture of total 100 µl was incubated at 37°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

Fifteen pmoles of the decapped mRNA and 150 pmoles of the DNA-RNA chimeric oligonucleotide synthesized in above REFERENTIAL EXAMPLE were dissolved in a 100 µl solution containing 75 mM Tris-HCl (pH 7.5), 0.1 mM ATP, 10 mM MgCl₂, 5 mM dithiothreitol, and 10 µl of DMSO. After adding 100 units of T4 RNA ligase (Takara Shuzo), the reaction mixture was incubated at 16°C for 16 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

Three µg of the DNA-RNA chimeric oligonucleotide-ligated mRNA and 1.5 µg of a pKA1 vector primer possessing a tail of c.a. 60 (dT)s (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292) were dissolved in a solution containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, 1.25 mM dNTPs(dATP + dCTP + dGTP + dTTP). After adding 200 units of reverse transcriptase (Takara Shuzo), the reaction mixture of total 20 µl was incubated at 42°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in a solution containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 4 mM MgCl₂, 10 mM (NH₄)₂SO₄, and 50 µg/ml bovine serum albumin. Then the reaction mixture was added by 60 units of E.coli DNA ligase (Takara Shuzo) and incubated at 16°C for 16 hours. After adding 2 µl of 2 mM dNTPs, 4 units of E.coli DNA polymerase (Takara Shuzo) and 0.1 unit of E.coli RNase H (Takara Shuzo), the reaction mixture was incubated at 12°C for one hour and then at 22°C for one hour.

Mixing 10 µl of the reaction mixture with 100 µl of E.coli HB101 competent cells, the transformation of HB101 was carried out according to the conventional method. The transformants grew on an ampicillin-containing LB agar plate. Twelve colonies generated on the plate were picked up, suspended in 3 ml of an ampicillin-containing LB medium, and incubated overnight. The plasmids were isolated from the cultured cells by the alkaline lysis method. The isolated plasmids were digested with EcoRI and NotI and analyzed on agarose gel electrophoresis. Each plasmid contained a cDNA insert of c.a. 600 bp. The sequence of the cDNA terminal was determined by the dideoxy method. The results showed that each cDNA possessed the sequence starting from the cap site of the model mRNA downstream of the synthetic oligonucleotide-derived sequence including an Ec-

oRI site.

EXAMPLE 2

Five μ g of rabbit globin mRNA (BRL) was dephosphorylated and decapped by the same way as described in EXAMPLE 1. Thirty pmoles of the obtained mRNA and 3 nmoles of the DNA-RNA chimeric oligonucleotide synthesized in REFERENCE EXAMPLE were dissolved in a solution containing 50 mM Tris-HCl (pH 7.5), 0.5 mM ATP, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, and 25 % polyethylene glycol. After adding 50 units of T4 RNA ligase (Takara Shuzo), the reaction mixture of total 30 μ l was incubated at 20°C for 12 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

After the reaction product was annealed with a pKA1 vector primer, cDNA synthesis and transformation of E.coli by the resultant cDNA vector were carried out under the same condition as described in EXAMPLE 1. Twelve colonies were randomly picked up from the agar plate. After incubation in the medium, the plasmid was isolated from the cultured cells. Since a rabbit globin cDNA has one EcoRI site at position c.a. 400 from the cap site, the full-length cDNA encoding rabbit globin should produce the fragment of 400 bp by EcoRI digestion. In fact, the double digestion of every plasmid with EcoRI and NotI produced two bands including c.a. 400 bp and c.a. 250 bp originated from the cDNA and one band corresponding to the vector. Those results suggested that every plasmid contained the full-length cDNA for globin. In order to confirm this, the 5' terminal sequences were determined. The results showed that each cDNA possessed the sequence starting from the cap site of the rabbit globin mRNA downstream of the synthetic oligonucleotide-derived sequence including an EcoRI site. The starting material, rabbit globin mRNA, is composed of alpha-globin mRNA and beta-globin mRNA. Out of 12 clones analyzed, 4 clones contained the full-length cDNA encoding alpha-globin and the remaining 8 clones contained that encoding beta-globin.

EXAMPLE 3

Using a DNA oligonucleotide prepared in REFERENCE EXAMPLE 3 instead of a DNA-RNA chimeric oligonucleotide, cDNA synthesis was done in the same way as described in above example. The number of transformants decreased by one tenth, but the sequence analysis of 5' end of cDNA isolated from the transformants showed that every clone contained the full-length cDNA encoding rabbit alpha- or beta-globin.

EXAMPLE 4

Three μ g of the DNA-RNA chimeric oligonucleotide-ligated rabbit globin mRNA prepared in EXAMPLE 2 was annealed with 1.5 μ g of the DNA fragment primer prepared in REFERENCE EXAMPLE, and then cDNA synthesis was carried out under the same condition as described in EXAMPLE 1. The product was digested with 50 units of EcoRI and 50 units of NotI, and ligated to 1 μ g of EcoRI,NotI-digested pKA1 (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292). The reaction mixture was used for transformation of E.coli. The plasmid was prepared from each transformant. The double digestion of every plasmid with EcoRI and NotI produced two bands including c.a. 400 bp and c.a. 250 bp on agarose gel electrophoresis. The 5' terminal sequences were determined by the dideoxy method. The results showed that each clone possessed the full-length cDNAs encoding the rabbit alpha- or beta-globin downstream of the synthetic oligonucleotide-derived sequence including an EcoRI site.

EXAMPLE 5

Ten μ g of poly(A)⁺RNA prepared from human fibrosarcoma cell line HT-1080 using the conventional method was dephosphorylated and then decapped in the same way as described in EXAMPLE 1. The obtained poly(A)⁺RNA and 3 nmoles of the DNA-RNA chimeric oligonucleotide synthesized in REFERENCE EXAMPLE were dissolved in a solution containing 50 mM Tris-HCl (pH 7.5), 0.5 mM ATP, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, and 25 % polyethylene glycol. After adding 50 units of T4 RNA ligase (Takara Shuzo), the reaction mixture of total 30 μ l was incubated at 20°C for 12 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

After the reaction product was annealed with the PKA1 vector primer, cDNA synthesis and transformation of E.coli by the resultant cDNA vector were carried out under the same condition as described in EXAMPLE 1. Six hundred colonies were randomly picked up from the agar plate. After incubation in the medium, the plasmids were isolated from the cultured cells. Determining the 5' terminal sequence of each cDNA clone, we found out that many cDNAs encoded proteins identical to known human proteins. In some genes encoding these proteins, their transcription initiation sites on the genomic sequence have already been determined, so that we compared it to the 5' terminal sequence of the obtained cDNA. In the case of abundant housekeeping proteins such as elongation factor 1-alpha, various ribosomal proteins and

so on, the 5' terminal sequence of the cDNA agreed with the sequence of the transcription initiation site. These results mean that the cDNA is full-length. Nine out of 10 clones encoding elongation factor 1-alpha contained the full-length cDNA (starting from 5'-CTTTTCGCAA...).

INDUSTRIAL AVAILABILITY

The invention enables sure synthesis of a full-length cDNA. Since a cDNA clone obtained from the cDNA library prepared according to the present method is sure to contain the whole information on the primary structure of the protein, the obtained clone can be used immediately to produce the encoded protein. Therefore, the present invention is useful for production of useful proteins by genetic engineering.

Claims

1. A process for producing an intermediate for the synthesis of a full-length cDNA, which comprises the steps of treating mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped degraded mRNA; decapping from the 5' end of a capped intact mRNA; and ligating either a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide represented by the following general formula [I] to the 5' end phosphate group formed in the above step by the action of T4 RNA ligase, thereby selectively adding either the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide having an arbitrary sequence to the 5' end of the intact mRNA.

$$5'\text{-dN1-dN2}\cdots\text{dNm-N1-N2}\cdots\text{Nn-3'} \quad [\text{I}]$$
 (wherein dN represents a deoxyribonucleotide selected from dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from AMP, CMP, GMP and UMP; "-" represents a phosphoric ester linkage; m represents an integer of 1 or more; and n represents an integer of 0 or more.)
2. The process according to claim 1, wherein said DNA oligonucleotide or said DNA-RNA chimeric oligonucleotide represented by the formula [I] contains at least one recognition site of a restriction enzyme RE1 which does not cleave mRNA-cDNA hybrid which is formed in a subsequent step.
3. A process for synthesizing a full-length cDNA, which comprises linking a double-stranded DNA primer having a dT tail by annealing to a poly(A) tail of the 3' end of the intact mRNA

having the DNA oligonucleotide or the DNA-RNA chimeric nucleotide added at the 5' end produced by the process according to claim 1, and then synthesizing the first strand cDNA complementary to the intact mRNA by a reverse transcriptase.

4. A process for synthesizing a full-length cDNA, which comprises linking a double-stranded DNA primer having a dT tail by annealing to a poly(A) tail of the 3' end of the intact mRNA having the DNA oligonucleotide or the DNA-RNA chimeric nucleotide added at the 5' end produced by the process according to claim 2, and then synthesizing the first strand cDNA complementary to the intact mRNA by a reverse transcriptase.
5. A process for preparing a recombinant vector containing a full-length cDNA, which comprises the steps of digesting with the restriction enzyme RE1 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by the process according to claim 4 wherein said double-stranded DNA primer is a vector primer containing at least one recognition site of said restriction enzyme RE1; circularizing the resultant material by self-ligation; and converting the RNA in the circularized recombinant vector to DNA.
6. A process for preparing a recombinant vector containing a full-length cDNA, which comprises the steps of digesting with the restriction enzyme RE1 and a restriction enzyme RE2 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by the process according to claim 4 wherein said double-stranded DNA primer is a vector primer containing at least one recognition site of said restriction enzyme RE2; recombining the resultant material with a vector having recognition sites of restriction enzymes RE1 and RE2 at its ends, respectively, by ligation; and converting the RNA in the circularized recombinant vector to DNA.
7. A DNA-RNA chimeric oligonucleotide.

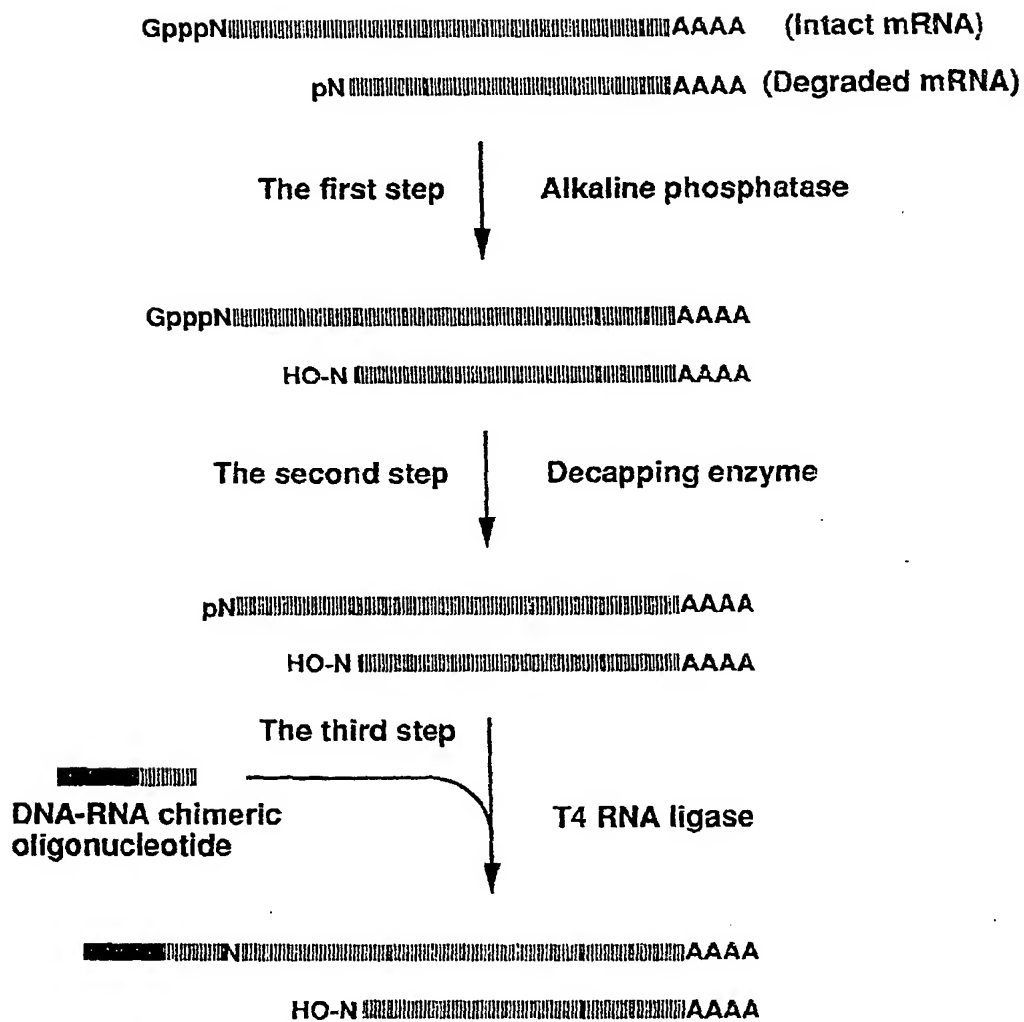


Fig.1

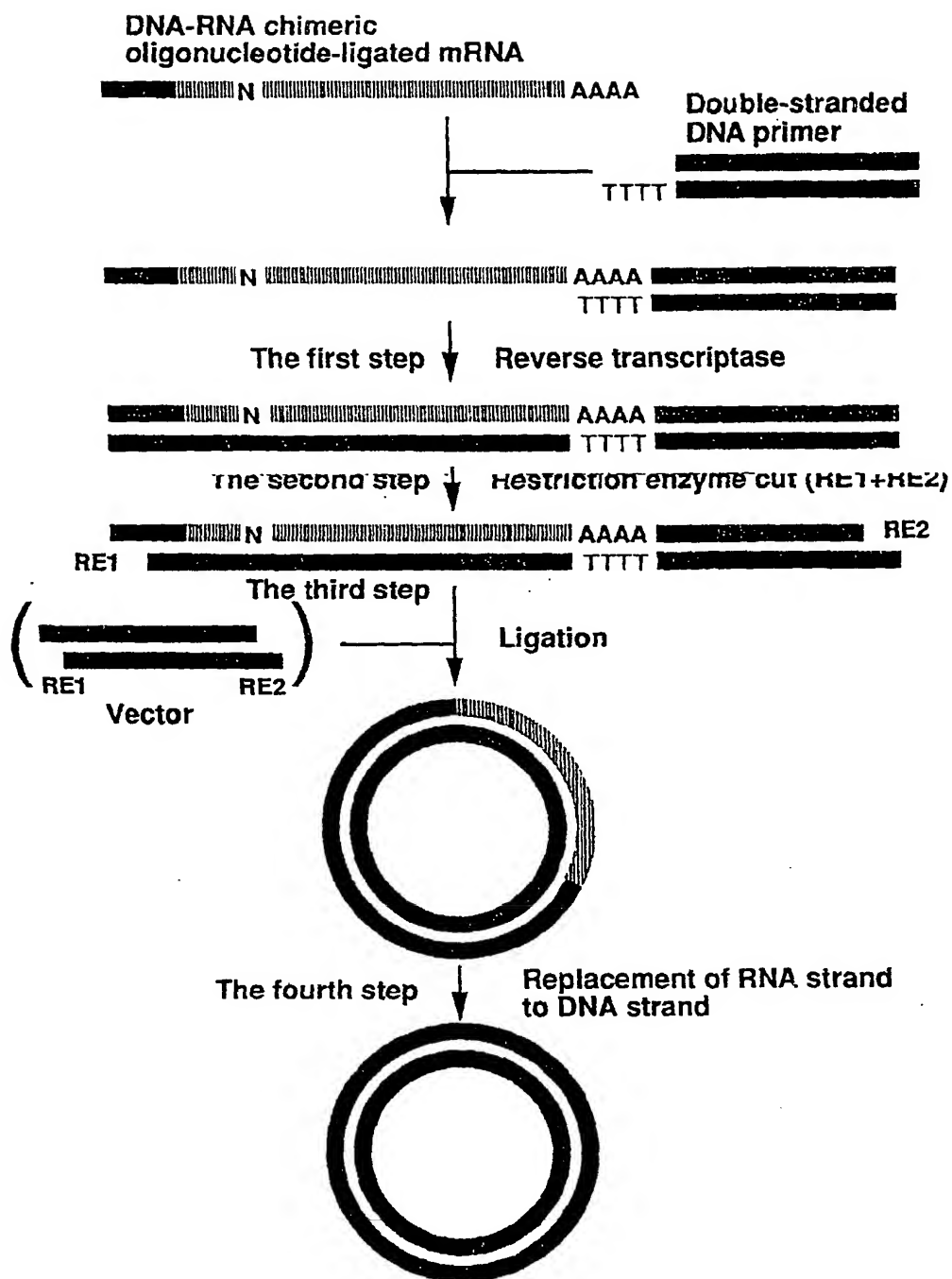


Fig.2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP93/01359

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl⁵ C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl⁵ C12N15/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS PREVIEWS, CAS ONLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Gene, Vol. 68, 1988, Rutledge RG et. al. "Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors" p. 151-158	1-7

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

December 21, 1993 (21. 12. 93)

Date of mailing of the international search report

February 8, 1994 (08. 02. 94)

Name and mailing address of the ISA/

Japanese Patent Office

Facsimile No.

Authorized officer

Telephone No.